

Gesdoc: un prototipo de software para repositorios de información digital

Piedad Garrido

Dept. de Informática e Ingeniería de
Sistemas
EUPIT Ingeniería T. Informática
Univ. de Zaragoza
44003 Teruel
piedad@unizar.es

Jesús Tramullas

Dept. de CC. de la Documentación
Facultad de Filosofía y Letras
Universidad de Zaragoza
Campus San Francisco
50009 Zaragoza
jesus@tramullas.com

José Romero Lafuente

Diputación Provincial de Teruel
EUPIT Ingeniería T. Informática
Univ. de Zaragoza
44003 Teruel
JRomero@dpteruel.es

Resumen

Este trabajo presenta GesDoc, una aplicación ad-hoc, de código abierto, para la gestión en línea de recursos digitales heterogéneos, creado en virtud de un proyecto de colaboración entre investigadores de los departamentos de Ciencias de la Documentación e Informática e Ingeniería de Sistemas, ambos de la Universidad de Zaragoza.

Se explican sus características y arquitectura, y se hace una descripción de la aplicación y de los servicios que ofrece. En el desarrollo de la aplicación ha desempeñado un papel fundamental la total integración de herramientas y estándares establecidos en el campo de las aplicaciones informáticas de gestión documental, como bases de datos ISIS, MARC21, XML y JAVA.

1. Planteamiento

El objetivo principal de este proyecto ha sido desarrollar una “aplicación distribuida”, que permita realizar operaciones básicas con bases de datos documentales, que almacenen recursos digitales heterogéneos.

Esta herramienta debe permitir a los usuarios realizar las operaciones más básicas como la creación de una base de datos y sus correspondientes campos y subcampos, la actualización y/o borrado del contenido, la importación y/o exportación de la información almacenada en diferentes formatos, etc. Se ha intentado aprovechar todo el potencial del motor de búsqueda que incorpora el software

documental. De esta forma, se pretende superar la limitación del modelo relacional de datos, donde el buscador es ajeno al repositorio de información y tiene que ser construido con un lenguaje de programación de alto nivel existente.

La idea de la construcción de dicha herramienta se plantea debido a que la cantidad de recursos digitales generados por la comunidad académica e investigadora se encuentra siempre en continuo crecimiento: publicaciones periódicas, tesis, directorios temáticos especializados, información audiovisual. Esta información suele ser buscada y almacenada por los investigadores o grupo de investigadores de forma individualizada, lo que conlleva problemas de duplicidad y redundancia a la hora de compartir información y documentación sobre determinados temas, y, lo que es más importante, formas muy dispares de referenciar cada uno de estos ítems informativos.

Al tratarse de un grupo emergente de investigadores, la puesta en marcha de un proyecto de este tipo resulta factible y a la vez muy recomendable, ya que ayudará a organizar la información de un determinado tipo de publicaciones de un departamento, empresa u organismo público, atendiendo a criterios similares de selección de información, organización y recuperación, evitando los problemas indicados en el párrafo anterior.

Para poder llevar un control de los recursos digitales y/o documentación que se generan en una determinada organización, en concreto en el caso que nos ocupa, la aplicación deberá permitir:

- El establecimiento de niveles jerárquicos de acceso y uso de la aplicación de las bases de

datos documentales, y de los documentos incluidos en las mismas.

- La creación de bases de datos documentales.
- La modificación de las estructuras de bases documentales existentes.
- La actualización de los contenidos de las bases de datos documentales, mediante la introducción, modificación y eliminación de los registros, y de los documentos asociados.
- El intercambio de datos con otros sistemas de gestión de bases de datos documentales. La aplicación deberá permitir la importación o exportación de registros en el formato de intercambio ISO 2709 MARC[8]. También deberá poder facilitar la exportación de dichos registros en XML. Para ello se utilizará el esquema desarrollado por *Network Development an MARC Standard Office de la Library of Congress*, que permite trabajar en XML con datos en formato MARC [10].

Por último comentar, que la aplicación está terminada, se encuentra en un período de pruebas con usuarios noveles en la materia para así poder llevar a cabo, a posteriori, un estudio de su funcionalidad y usabilidad. Se espera que empiece a estar operativa el curso académico que viene.

2. Arquitectura de Gesdoc

Previamente al desarrollo de la aplicación se ha llevado a cabo un estudio de las aplicaciones y herramientas existentes para manejar este tipo de bases de datos: JavaISIS, WWWISIS, GenISIS, Elvis, WinISIS, Utilitarios CISIS, etc. [9,13,14] En el marco del mismo, se analizaron los siguientes aspectos:

1. Organización de la información en las bases de datos documentales.
2. Soporte para trabajo en grupo.
3. Formatos de intercambio de datos.
4. Plataformas que soportan las aplicaciones.
5. Aspectos de seguridad.

Tras la toma en consideración de los factores indicados, se tomó una decisión con respecto al entorno de programación y las tecnologías más adecuadas para el desarrollo del proyecto.

Como se puede observar en la figura 1, la aplicación consta de una aplicación servidora, una

aplicación cliente y, un gestor de bases documental con información formateada preparada para soportar el formato MARC21[7], que implementa la norma ISO 2709.

2.1. Aplicación cliente

El cliente deberá ser capaz de funcionar correctamente en cualquier PC con un sistema operativo que tenga instalada la máquina virtual de Java.

La aplicación cliente se conecta a un servidor web, para acceder a las diferentes bases de datos de dicho servidor, bajo un sistema basado en ventanas que interactuará con las bases de datos ubicadas en diferentes servidores, a través del protocolo http.

2.2. Aplicación servidora

El servidor podrá funcionar en un servidor o PC en el que se encuentre instalado un contenedor de servlets, como puede ser Tomcat, o un servidor web, capaz de invocar servlets de Java.

La parte de la aplicación del servidor, deberá residir en un servidor web, capaz de trabajar con el protocolo http.

Sin embargo, en la arquitectura de la aplicación se observa que el lenguaje de bajo nivel para el intercambio de información estructurada elegido es XML. Lo que aporta una gran estabilidad al proyecto ya que únicamente habría que implementar la interfaz para generar nuevas formas de acceso a la base de datos. Por ejemplo, desde terminales móviles, PDA's, etc.

2.3. MARC & XML

La relación entre MARC y XML data de 1992, con la inclusión del campo 856 (*electronic location and access*) propuesta por la *Library of Congress*, LOC. En 1995 surgió una segunda iniciativa, también por parte de la LOC, que consistía en adaptar MARC a SGML y viceversa. Fruto de la misma, en enero de 1998 se hizo pública la DTD MARC SGML.

Actualmente, la *LOC Network Development and MARC Standards Office* ha publicado un esquema XML para codificar registros MARC21. Este esquema, resultado del trabajo de la LOC; OCLC, RLG y, revisado por la *Nacional Library*

of Medicine y la National Library of Canada, ha sustituido a la iniciativa anteriormente derivada de la DTD SGML.

A raíz de este esquema, presentado en la figura 2, se han puesto en marcha numerosas iniciativas y herramientas:

- MARC4j

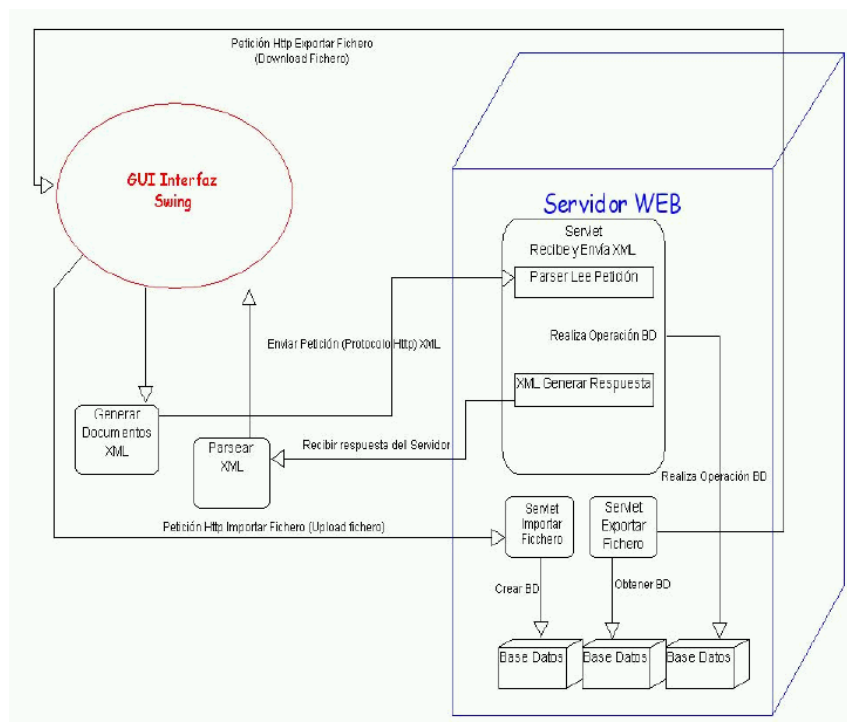


Figura 1. Arquitectura de la aplicación desarrollada.

- Hojas de estilo XSLT para convertir MARC XML a MODS, Dublin Core, OAI MARC y HTML.
- MRC2XML
- XMLMARC
- BIBLIOML
- MODS

El utilitario con el que se ha trabajado en el proyecto ha sido MARC4j [4], ya que permite exportar e importar el contenido de las bases de datos en MARCXML. Se utilizará el esquema de MARCXML para transmitir la base de datos del cliente al servidor, aunque se hará uso de un

parser propio, diseñado exclusivamente para este proyecto y que permitirá trabajar internamente con la base de datos.

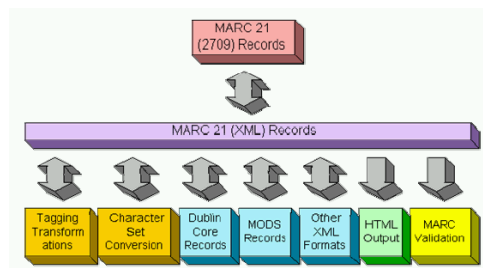


Figura 2. Esquema de conversión.

Para terminar este apartado hay que enfatizar que se ha puesto especial cuidado en aspectos como la representación de la información, el desarrollo de una aplicación servidora robusta y multiplataforma, ambos con la finalidad de que la aplicación cliente soporte el mínimo de información a procesar, con el objetivo de que estas bases de datos documentales, que almacenan recursos digitales heterogéneos, puedan ser consultadas y gestionadas tanto desde un PC de sobremesa como desde un PDA o un dispositivo móvil.

3. Bases de datos ISIS

El software Micro CDS/ISIS, (*Computerized Documentation System - Integrated Set for Information System*), también conocido como MicroISIS, es un sistema generalizado de almacenamiento y recuperación de información, diseñado especialmente para el manejo computarizado de bases de datos documentales.

Etapas de su evolución [1]:

ISIS:

La Organización Internacional del Trabajo (OIT) había creado para su uso interno el "Integrated Set of Information Systems" – ISIS.

CDS/ISIS:

Para el sistema de documentación de la UNESCO (*Computerized Documentation System, CDS*), Del Bigio propuso basarse en el ISIS de la OIT.

El nuevo CDS fue programado en lenguaje PL1 y denominado CDS/ISIS

MicroISIS:

Después de su primera presentación mundial en la Reunión de Usuarios del Sistema CDS/ISIS para América Latina y el Caribe realizado en la CNEA (Octubre 1985), el Mini-micro CDS/ISIS, fue entregado a los primeros usuarios en un curso realizado en la sede de la UNESCO en París en diciembre de 1985.

WinISIS

Es la versión para el sistema operativo Windows de este sistema gestor documental.

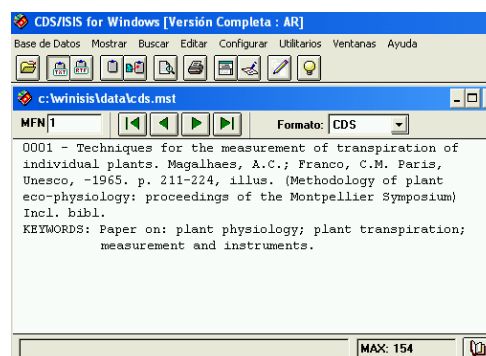


Figura 3. Interfaz de WinISIS

OpenISIS [15]:

Es una librería JAVA de código abierto (*Open Source*) que si bien está en etapa experimental es una excelente opción a la hora de realizar cualquier tipo de desarrollo con bases de datos ISIS que funcionen sobre otro sistema operativo, que no sea Microsoft Windows. Implementa UNICODE, por lo que permite desplegar texto con cualquier set de caracteres.

Este ha sido el repositorio de información integrado en el proyecto. La aplicación crea una base de datos ISIS, independiente del sistema operativo, lo que resulta muy útil ya que la mayoría de los resultados y software freeware que se va creando en torno a este gestor documental han sido desarrollados para su funcionamiento bajo Windows, y con herramientas de las que el usuario final que quiera modificar código tiene que tener licencia.

4. Descripción de la aplicación

Al arrancar la aplicación, el usuario se encontrará ante la interfaz mostrada en la figura 3, que le obligará a validarse en un servidor determinado para acceder a la aplicación. Es importante señalar que cada servidor de la aplicación tiene distintos usuarios. Es decir, una misma persona puede acceder a distintos servidores como usuario en ellos y sus claves de acceso pueden ser distintas en cada servidor.

4.1. Módulo principal

Desde el menú de opciones, señalado como 1 en la figura 3, se pueden ejecutar todas y cada una de las operaciones a realizar por la aplicación Gesdoc. Este menú varía según el perfil de usuario.

Si el usuario es:

- Administrador del dominio.
- Administrador de la base de datos.
- Lector

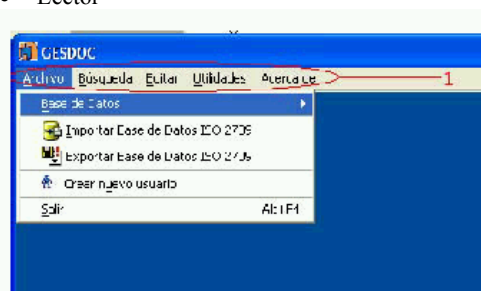


Figura 4. Módulo principal

4.2. Módulo de búsqueda

Una vez seleccionada la base de datos documental (ver figura 4) con la que se desea trabajar, se selecciona el módulo de búsqueda.

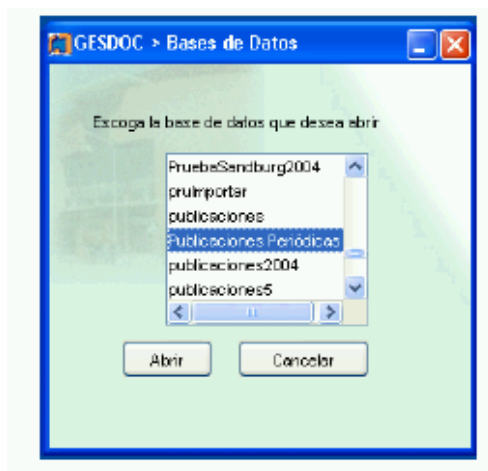


Figura 5. Abrir Base de Datos

La pantalla de búsqueda, puede ser vista por cualquier usuario que tenga acceso a la base de

datos, y permite buscar un contenido determinado en los registros en base a la combinación de distintos campos haciendo uso de operadores booleanos y eliminando toda palabra vacía que no aporte ningún significado concreto.

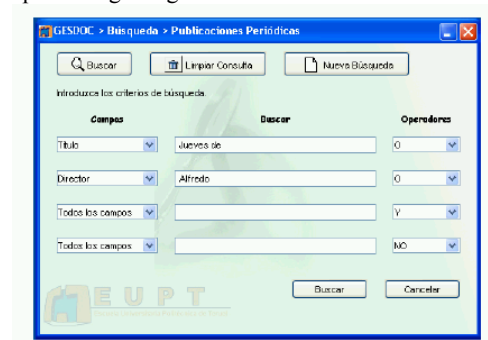


Figura 6. Módulo de búsqueda Gesdoc

Integrando en una única interfaz la búsqueda guiada y experta que proporciona ISIS, y que para un usuario inexperto que se enfrente a la consulta de una base de datos documental le resultará muy complicado.

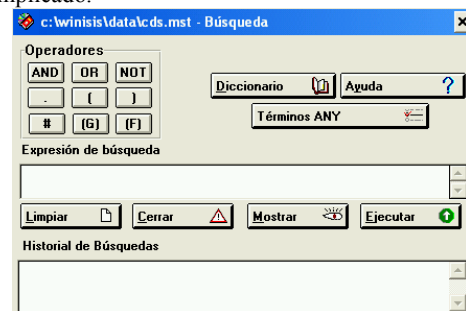


Figura 7. Búsqueda Experta WinISIS

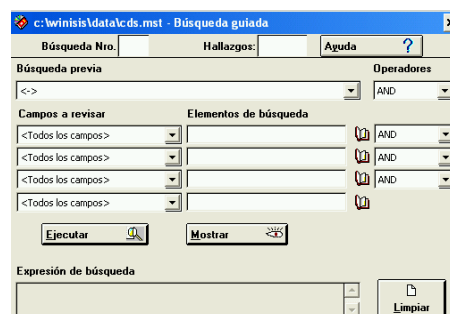


Figura 8. Búsqueda Guiada WinISIS

Siendo mucho más intuitiva y amigable la interfaz de búsqueda creada en Gesdoc (ver figura 6).

4.3. Módulo de edición

En este apartado se estudiará la interfaz que permite a los usuarios con al menos derechos de editor sobre la base de datos, modificar el contenido de los registros de la misma.

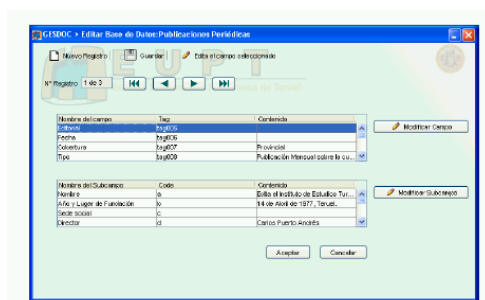


Figura 9. Módulo de edición Gesdoc

Como se observa en la figura anterior se brinda acceso rápido a operaciones tales como: Nuevo Registro/Guardar/Editar el campo seleccionado. Para desplazarnos de forma eficaz, existe una barra de navegación análoga a otras barras de navegación aparecidas en otras opciones del programa, con la finalidad de mantener una coherencia y evitar que el usuario se despiste.

En la parte central de la pantalla, se muestran dos tablas, que pertenecen a los campos y subcampos de la selección realizada previamente y que muestran el contenido del registro actual seleccionado.

Tarea totalmente guiada, en contraposición, por ejemplo al uso de WinISIS donde hay que etiquetar todos los datos de forma manual, teniendo que tener un conocimiento profundo del formato de representación utilizado.

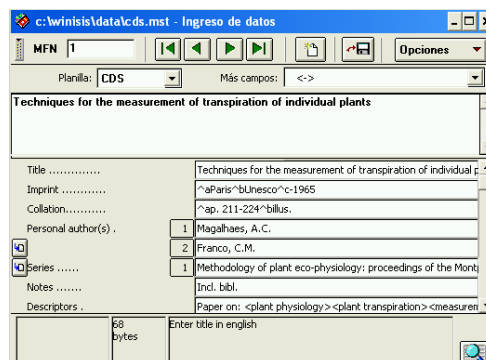


Figura 10. Entrada de datos WinISIS

4.4. Módulo de creación

Lo único que se le exige a un usuario para dar de alta una nueva base de datos es que sea administrador del dominio. El proceso de creación de la base de datos, en contraposición con WinISIS, es tan sencillo como introducir un nombre y una breve descripción para la base de datos.



Figura 11. Crear base de datos (Paso 1)

A continuación, se tiene que pasar a crear el diseño de la base de datos documental. Crear los campos y subcampos de una manera mucho más sencilla e intuitiva que con el WinISIS original.

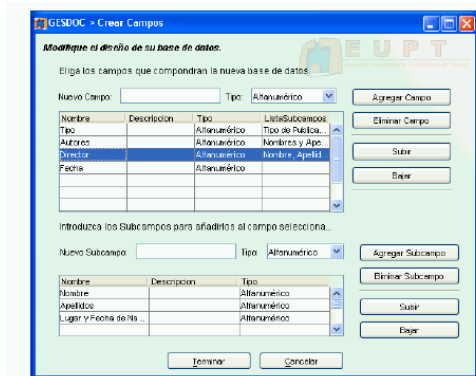


Figura 12. Creación de Campos (Paso 2)

4.5. Módulo de Control de acceso

Las tareas de control de acceso a una base de datos ubicada en un determinado servidor, sólo pueden ser realizadas por un administrador de la base de datos. Desde aquí se pueden realizar operaciones tales como:

- Introducir un nuevo usuario en el dominio.
- Modificar el control de acceso dado a un usuario.

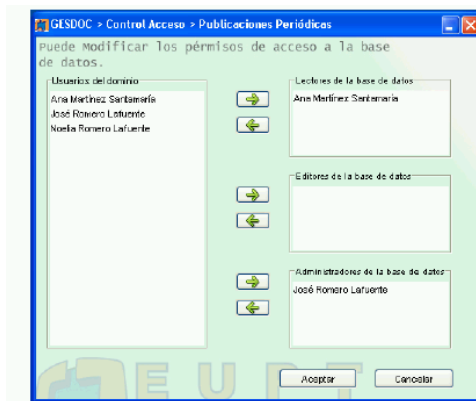


Figura 13. Modificar control de acceso

4.6. Módulo de Exportación/Importación

Ambas son tareas que al igual que ocurre en el módulo anterior sólo pueden ser llevadas a cabo por el administrador de la base de datos. La aplicación puede exportar sus datos en 2 formatos:

6. ISO 2709 (MARC21)

7. MARCXML [17]

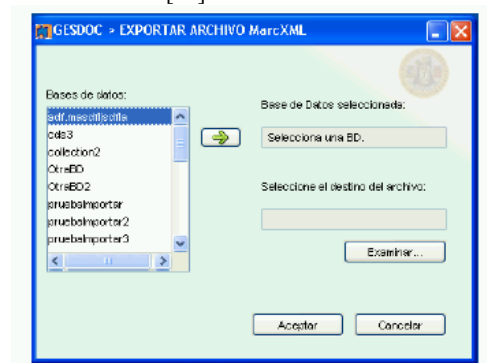


Figura 14. Exportar Base de datos

En cuanto a la tarea de importación, la aplicación también puede importar los datos en los dos formatos anteriormente comentados para la exportación.

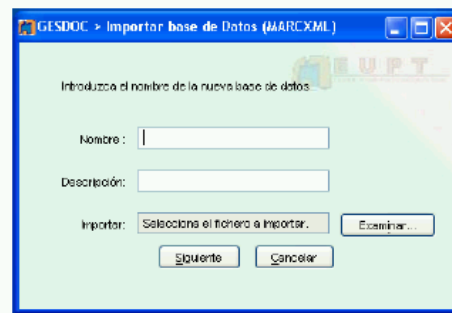


Figura 15. Importar Base de Datos

Obviamente para importar una base de datos de un fichero proporcionado, éste debe cumplir con el formato correcto, y deberá ser acompañado por los parámetros: nombre y descripción de la base de datos.

5. Conclusiones

La primera de las conclusiones a comentar es que ha sido un trabajo muy fructífero debido al intercambio de conocimientos entre integrantes de diferentes disciplinas.

Sobretudo el departamento de Informática e Ingeniería de Sistemas ha encontrado grandes dificultades a la hora de enfrentarse a un tema desconocido como la descripción del contenido documental y el uso alternativo de una base de datos documental en contraposición a un sistema gestor de bases de datos relacional.

Una vez finalizado el proyecto, queda bastante claro qué información que se tenía anteriormente almacenada en pequeñas bases de datos relacionales de forma local, tiene ahora un gran valor añadido, al cumplir con un formato de representación de la información acorde a un estándar. Además, es viable intercambiar información con otros centros de documentación, bibliotecas, etc. a nivel nacional e internacional.

Que no es sino un objetivo fundamental de toda buena tarea de selección y recopilación de información llevada a cabo por todo buen investigador.

Surgiendo así fuentes de información nuevas, ya que gran parte de las fuentes primarias de información pueden ser hoy en día encontradas en la web en páginas particulares, web departamentales, directorios temáticos especializados, etc.

También se consiguen subsanar problemas de redundancia y duplicidad de información, que de hecho se daban entre investigadores de la misma área.

Para finalizar comentar, aunque se ha hecho previamente, que la aplicación resulta muy útil y es bastante novedosa, ya que para tener instalado un servidor de base de datos documental ISIS, se debe hacer uso del sistema operativo Windows. Actualmente, no existe otra alternativa ya que el motor de base de datos Malet, creado por la iniciativa de software libre de la familia CDS/ISIS está en pleno desarrollo. De ahí que con esta arquitectura software se haya conseguido uno de los objetivos tan perseguidos en los desarrollos en los que participa esta base de datos documental, que no es otro que el software no sea sólo gratuito y se disponga de un ejecutable, sino que también sea abierto.

Referencias

- [1] Manual para instructores de WinISIS.
<http://www.cnea.gov.ar/cac/ci/isis/manual/>
- [2] API Servlet
<http://java.sun.com/products/servlet>
- [3] Web Oficial Proyecto JDOM
<http://www.jdom.org/>
- [4] Web Oficial Proyecto MARC4J
<http://marc4j.tigris.org/>
- [5] Web Oficial de Bireme
<http://www.bireme.br/>
- [6] Extensible Markup Language W3C
<http://www.w3.org/XML/>
- [7] Formatos MARC: MARC21
<http://www.loc.gov/marc/>
- [8] ISO 2709
<http://www.cindoc.csic.es/isis/ap-b.htm>
- [9] JAVAISIS
<http://web.tiscali.it/javaisis/>
- [10] Proyecto MODS
<http://www.loc.gov/standards/mods/>
- [11] MARC21
<http://www.loc.gov/marc/bibliographic/ecbdhome.html>
- [12] UNESCO/ISIS
<http://www.unesco.org/isis>
- [13] WWWIsis
<http://productos.bvsalud.org/product.php?id=wwwis&lang=en>
- [14] Utilitarios CISIS
<http://productos.bvsalud.org/product.php?id=cisis&lang=es>
- [15] OpenIsis
<http://openisis.org/Doc/>
- [16] Swing for Java
<http://java.sun.com/docs/books/tutorial/uiswing/>
- [17] Iniciativa XMLMarc
<http://xmlmarc.stanford.edu/>
- [18] BiblioML
<http://www.biblioml.org/>
- [19] P.Garrido, F.J. Martínez. MPEG-1 Layer 3 y WINISIS: La tecnología al servicio de la Documentación". Actas de las I Jornadas de Bibliotecas Digitales, anexas a las V Jornadas de Ingeniería de Software / V Jornadas de Bases de Datos (Valladolid, noviembre de 2000) Departamento de Informática de la Universidad de Valladolid, pp. 299-302